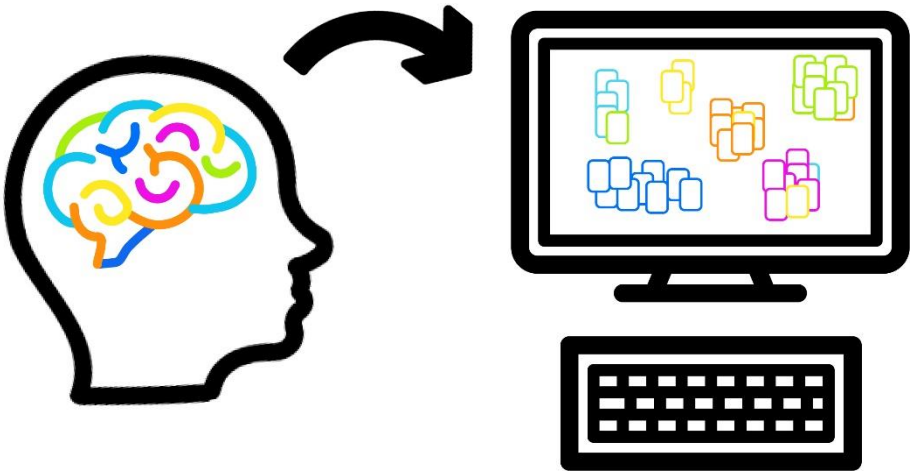
Website Navigation Structures:

Eliciting Mental Models of University Students using Card Sorting and the Perceived Usability



Linda E. M. Renes

Faculty of Behavioural, Management and Social Sciences, University of Twente

Bachelor Thesis

First Supervisor: Marlise Westerhof

Second Supervisor: Dr. Martin Schmettow

23-01-2023

Abstract

The aim of this research is to elicit the mental model of university students regarding the educational page of university websites and to determine whether the perceived usability rises if a website matches this mental model. Since there is little knowledge yet about the mental model of university students in this field, an open, remote card sort study took place. In this study, a structure of five clusters with four subclusters is found. Also, four ambiguity groups are found and interpreted. By comparing this cluster structure to already existing websites, two sets of three tasks are created; one for tudelft.nl/en/education where the information goals are in the right place according to the mental model, and one set for tue.nl/en/education where the information goals are not in the correct place according to the mental model. After performing these tasks, participants filled out the System Usability Scale (SUS) for each website. The SUS is a valid tool to measure the perceived usability of systems and works for a small sample. The results of the SUS show, contrary to the expectations, that the perceived usability of the website where the information is "in the wrong place" is higher. When the participants are asked directly which website they prefer, the majority indicates the website where the information goals are "in the right place." Since the results of this research indicate that no relation is found between the mental model and the perceived usability, more research is necessary in this field to determine whether there are relations or are no relations between these two subjects.

Keywords: Mental model, card sorting, university website, information architecture, perceived usability

Table of Contents

Introduction

Nowadays, the usage of the internet and websites is the most used way when looking for information (Case & Given, 2016). It has come to the point where a lot of information for all different kinds of companies is only to be found online; therefore, this information should be easily accessible. An example of a website that contains a lot of information for a lot of different people is a university website (e.g. students, employees, high scholars, researchers, etc.). Students, for example, need to find information about their study, teachers, and possible master tracks. The information structure of a website is created by people with knowledge of the field in which a website provides information. This knowledge creates a bias towards where experts think end-users will look for information (Bussolon, 2009). Even though this inconsistency exists, the information should be easy to find for all students. Therefore, this paper aims to research what a logical structure of information for university students on a university website is and to determine if this indeed makes the website easier to use for students.

Information Architecture

Website navigation is the way people look for information on a website. This includes, for example, how people scan through websites and on which items they click in the menu of a website (Madejska & Persson, 2019). Where information is placed on a website is also known as information architecture (Morville & Rosenfeld, 2006). An information architecture that is properly structured can increase the organization of information, thereby supporting the usability and findability of information in on websites (Morville & Rosenfeld, 2006). How usable a user experiences the website, conscious as well as unconscious, is called perceived usability (Lewis et al., 2015). Higher perceived usability has an immediate positive effect on the user's satisfaction and trust in the website (Flavián et al., 2006). It is explained, that to increase the perceived usability of a website, simplicity, interactivity, and perceived logic are essential (Madejska & Persson, 2019; Roshan & Ahmadi, 2022; Schall, 2014). So, to have a website that is satisfactory to use it needs to have an information structure where information is easy to find.

Mental Model

A way to create an information architecture where information is easy to find, is to align the structure with the mental model of the user. Firstly, in this paragraph, it will be explained what a mental model is. In the following paragraph it will be explained how this connects to the information architecture. Rook (2013) defined a mental model as: "a concentrated, personally constructed, internal conception, of external phenomena, or experience, that affect

how a person acts" (p. 42). In a more simplified version, Westbrook (2019) says that the mental model is an expectation of how something works and is subconsciously constructed. Jones et al. (2011) make the addition that a mental model should be viewed as a computational structure, meaning that a mental model is created in the working memory and then run like a computer. This gives a person the possibility to explore and test multiple options mentally before they act (Jones et al., 2011). Johnson-Laird (2010) explains that our "intuitive system" creates the mental model. Being part of this intuitive system means that it is our first instinct to handle according to the mental model and handling according this mental model is the fastest way to take effortless action (Johnson-Laird, 2010). Looking at the context for a website, this means, that it influences where someone would assume certain buttons to be placed, but also where certain information would be found.

Looking at the mental model of an end-user can be used to improve the information architecture of a certain website. If information is placed where a user expects it to be, it is easier for them to find. Several studies have shown that a website where the information architecture is compatible with the mental model of the user it has a higher perceived usability and satisfaction (Dillon & Turnbull, 2005; International Standards Organisation, 1998; Nielsen, 2010; Vinney, 2021). For an information architecture to be created based on a mental model, it needs to be known how the mental model of the end-user works concerning the presented information. However, since every user has a slightly different mental model, the information architecture should be created based on the average mental model of the end users (Nielsen, 2010). This means, that an information architecture for a satisfactory website should be in line with the average mental model of the end-user.

Card Sorting

A method that has proven itself to be effective when it comes to determining a mental model is card sorting. Card sorting is a research method that is used to understand how end users group information. Experts can use this to integrate the end users expectations into their ideas (Schmettow & Sommer, 2016). There are two types of card sorting, closed and open card sorting. In both types of card sorting the users are presented with a certain amount of cards from which they should create groups that are logical to them (Maguire, 2001). When using closed card sorting, category names are provided into which the cards should be divided. For open card sorting participants should formulate names and create categories themselves (Schmettow & Sommer, 2016). In this research an open card sort is used, therefore, when referring to card sorting, an open card sort is meant. When using card sorting, items will be put on the cards for

the card sorting task that contain information from which the mental model needs to be known. Card sorting can take place on pieces of paper handed to a participant, however, nowadays often a digital version is used where participants will receive online cards to sort (Bussolon et al., 2006). When participants have sorted the cards, an analysis can be performed to see which of the items were often put in similar categories (Maguire, 2001). The aim of performing this analysis is to find the average mental model of an end user (Bussolon et al., 2006). By creating this average mental model, a basis is created on which the most ideal navigation architecture can be built.

Card sorting studies

Card sorting has been used in many different researches to identify several ways to organise information. Often, these results show in practically integrated systems; for example, Cornell University used card sorting to organise the help topics on their library website (Faiks & Hyland, 2000). Also, Massachusetts Institute of Technology libraries conducted an open card sorting experiment about information from their website. Combined with the follow-up research, where participants had to point out in which of the five give categories they expected to find certain information, this resulted in a redesign of the website (Hennig, 2001). As these examples show, Schmettow and Sommer (2016) describe card sorting as a valuable method to gather input from users, to design an information structure. Both these example researches have the aim to increase the usability of their website by aligning the information architecture of their website to the mental model. Nakhimovsky et al. (2006) validated the method, card sorting, in a research about frequently asked questions, sorted by experts. Also, a research about redesigned operator stations of military flight simulators show an improvement between the original design and the redesign based on a card sort (Branaghan et al., 2011).

Contrary to the validations found by Nakhimovsky et al. and (2006) and Branaghan et al. (2011), Schmettow and Sommer (2016) did not find a relation between tasks compatible with the mental model and a better browsing performance. In the first part of their research Schmettow and Sommer (2016) let students from the University of Twente sort 69 cards with items from the municipality website of Enschede into stacks with a maximum of three levels. Following, it is checked for four other municipal websites whether the items are found there; this resulted in a set of 35 items present on all websites. For these 35 items a “mismatch” score is calculated for all items of all five websites, where a lower mismatch score means it is more in line with the mental model. The participants in the second part of the study performed five search tasks, one on each website. When looking at the results there was no relation found

between the browsing performance and the mismatch score (Schmettow & Sommer, 2016). Since there are results that do not find a relation between structures based on card sorting and browsing performance, as well as results validating the method of card sorting, it is interesting to look at the validation of card sorting results.

Perceived Website Usability

The aim of using a mental model as basis for the information architecture is to increase the usability of the website for the end user. Otter and Johnson (2000) explain that a complicated information structure creates a feeling of "lostness" and frustration. Lah et al., (2020) have also refuted claims that perceived usability would not be a useful construct. Before this, Diefenbach et al. (2014) identified perceived usability as a fundamental exponent of user experience. A common way to determine the perceived usability of a website is by the use of a standardized questionnaire. The usage of standardized questionnaires has proved to be valid and reliable in the measurement of perceived usability (Lah et al., 2020). Moreover, if the perceived usability of a website is higher a company (or university) will have a more trustworthy and overall more positive image for people looking up information on their website (Flavián et al., 2006).

Research Aims

Since more and more information is only available online, the importance of an information architecture that makes it easy for people to find information rises. On university websites, it is important that students can reach their information goals (e.g. finding contact details of a teacher or specifics about a course). Moreover, when the information architecture of a website is perceived better, the perceived usability of a website is higher. This has a positive effect on the public image of the university. The usability for university students can be enhanced by matching the information architecture of the website to the average mental model of university students. For university students, the information on the educational page is most relevant, since most information they need is found in this part of the website. To gain more information about the mental model of university students regarding information of the educational page of a university website, the aim of this research is: *"Eliciting the mental model of university students regarding information of a university website and to determine whether the perceived usability rises if a website matches this mental model"*. To reach this aim, the research questions that will be answered are: (1) *"What is the average mental model of university students regarding information on the educational page of a university website?"* and (2) *"To what extent does the perceived usability of a website increase for university students if the website matches the average mental model of a student?"*

Methods Card Sorting

Participants

In this study 45 participants took part. Participants were a convenience sample and were recruited either by being familiar to the researchers or through "Sona Systems"; a test subject-pool where psychology students can receive course credits. Participants who are acquaintances of the researchers received a link to the study via WhatsApp or E-mail. For the participants through Sona Systems, this link was provided in Sona Systems. Of the 45 participants, twelve were excluded based on having created categories with invalid names or creating one or two categories. This came to a final sample of 33 participants. Six of these identified as male, 26 as female, and one as other. Of the sample twelve participants were Dutch, sixteen were German, and five from another nationality. Sixteen participants were aged 18-20, ten were 20-22, six were 22-24, and one was 25 or older.

Selecting Items for Card Sorting Task

To find items that should be placed on the cards for the card sorting task, the educational part of eight Dutch university websites were analysed (Appendix A). These Universities were chosen because they are widespread over The Netherlands and have different focuses (e.g. not only technical universities). To find possible items, firstly, the website of the University of Twente (UT) was analysed since this website is most familiar to the researchers. Of this website menu items, headings, and other items that represent the information on the website were noted. After this, the researchers each checked some of the other university websites for additional items and whether the items from the UT website were found on this website as well. Additional items that were found on the other website, but not the UT website were also added to this list. The complete list of items can be found in Appendix B.

There were more items found than ideal for a card sorting task. The most optimal amount of cards is 30 – 40 for the method to have accurate results (Card Sorting, 2013). Moreover, the researchers doubted the relevance of some of the items for the educational page of a university website. To select which items should be placed in the task, it was decided that items that occur on more than half of the websites should be in the task; thus on at least five out of eight of the chosen websites. If the information was found on more than half of the websites, this information is considered to be relevant for students visiting this part of the website. An overview was created of the frequency on which the items were found on the eight websites. This resulted in the final list of 34 items for the cards. The items, including translations from Dutch to English, can be found in Appendix C.

Materials

For this study, the main material was a task created in the programme "kardSort". This task consisted of five parts; the informed consent, questionnaire, instructions, card sorting task, and thank you page. There was a questionnaire (Appendix E) to collect the demographic data of the participants regarding their age, gender, nationality, native language, and level of education. Additionally the question *"How often do you use University websites on average?"* was added with answering options varying between *"Daily"* to *"Less than 3 times a year".* Also, the familiarity with card sorting was collected by the question *"How familiar are you with Card Sorting?"* and the answering options varying from *"I don't know it"* to *"I used it as a research method myself".* The instructions (Appendix F) for the Card sorting task and usage of the programme were provided. The main task, the card sorting (Appendix G) consisted of 34 cards on the left and a blank space on the right where the categories were created.

Procedure

When clicking the link, the participants immediately arrive on the first page which is the informed consent (Appendix D). The informed consent explains that participation is voluntary and a participant can withdraw at any moment and their data is handled confidential. By selecting the "Next" button the participants agreed to informed consent. The demographic questionnaire followed the informed consent.

After the questionnaire, the instructions for the main card sorting task were provided. The participants were instructed to create several categories with meaningful names. These categories showed up on the right side of the screen. The 34 cards with items on them on the left side of the screen needed to be dragged into the correct categories (see Appendix G). An example of partly sorted cards can be found in Figure 1 The instructions told the participants they were unlimited in the amount of created categories. Moreover, when a participant was doubting in which category an item should be placed they should choose the best one. At the end of the task, when all the cards were sorted, the participant was thanked for their participation on an acknowledgment page (see Appendix G).

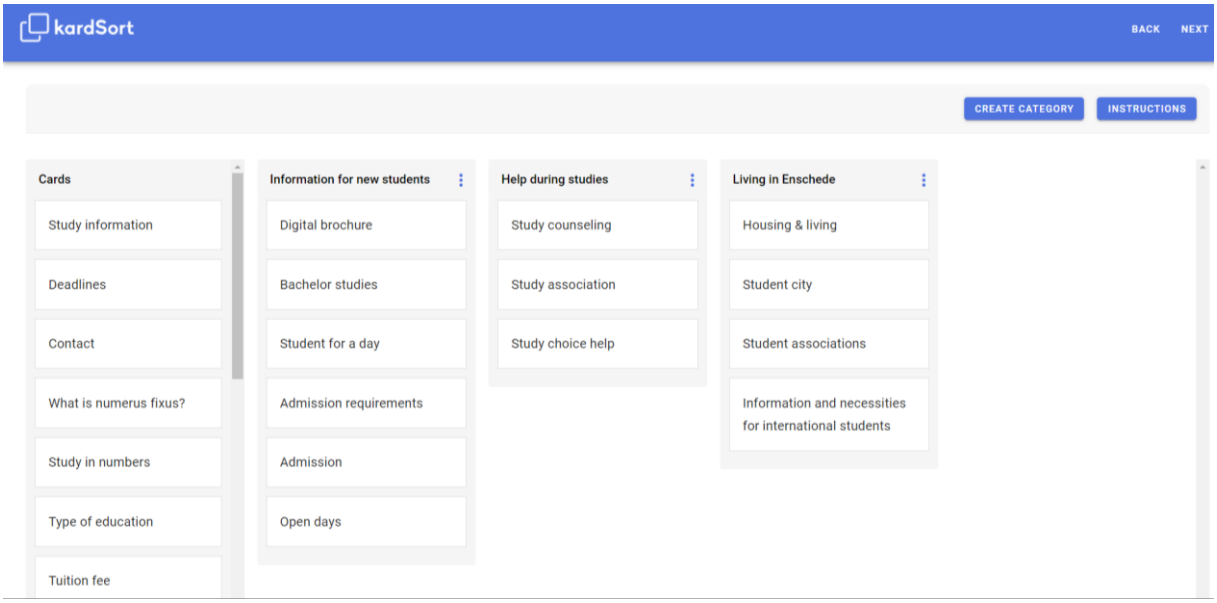
Data Analysis

Before the data was analysed, the results of the participants were observed to check if all participants created results usable in a cluster analysis. The entries were considered unusable when there were less than three categories created and when created labels were not meaningful (Righi et al., 2013). For example, participants who created only two categories with names like "general info" and "study info" were excluded. Moreover, when meaningless category names

or other names that caused suspicion (e.g. “X”, “I don’t know what this means”, or “3”) were created, these participants were excluded from the study if the items in the categories were not comparable with categories other participants created. In the end, twelve participants were excluded, which means there was a final sample of 33 participants. According to Tullis and Wood (2004), it is necessary to have at least 20-30 participants to have valid results for card sorting. Paul (2008) supports this and suggests card sorting studies even produce correct results with fewer participants. In R the demographic information of the dataset was analysed on age, gender, and nationality to create an overview of the population.

Figure 1

Example of the Card Sorting Task



Note. Part of the cards are sorted into categories as an example, these are not real results.

Heatmaps

To perform the data analysis of the card sorting task, the cleaned dataset was uploaded in the program SynCaps V3. The heatmap showed the results of the card sorting task, a similarity matrix, in a visual manner. This programme automatically created heatmaps from the data collected from kardSort (preview Figure 2). The heatmaps are created by ordering the items based on the semantic proximity, which is an advantage in comparison to the other analysis methods. This means that the items are reordered in a way that semantic clusters can be found along the diagonal forming the ‘hot patches’ on the diagonal axes (Hudson, 2005). A similarity matrix is dichotomous, thus it is shown in percentages how often each item was placed in the

same group as each of the other items (Wood & Wood, 2008). In the heatmaps a higher percentage (similarity score), thus higher semantic proximity is seen as a darker colour. This gave a reflection of how strong the mutual semantic proximity is between the individual items (Schmettow & Sommer, 2016). Where card sorting is a qualitative task, meaning the data is from observation and meant to identify specific problems (Budio, 2017); turning the data into similarity scores and a heatmap provides the possibility for quantitative analysis (Schmettow & Sommer, 2016).

Cluster analysis

Using the created heatmap, the mental model was analysed by observing which clusters were found. Clusters are groups of items that were often categorized together. The groups were recognized in a heatmap because they have darker colours and thus higher similarity scores. When items were placed in the same cluster, participants expected to find information about these items at the same place on a website. Clusters were considered conclusive when they had a minimal general similarity score of around 50%; thus at least half of the participants put these items together in a category. With a similarity score between 25% and 50% the scores were seen as inconclusive; thus there was doubt. Under 25% the similarity score was considered low and there was no reason to group those items together.

Naming clusters

Since the card sort was an open card sort, names for the found clusters or categories had also to be found. SynCaps V3 also provided an item-by-category matrix which showed 190 suggestions for categories. To find proper names the category labels were adjusted and merged (see Appendix H). For labels to be merged they had to either only differ in grammar (e.g. “study requirements” and “requirements for study”); have the same semantic meaning (e.g. “before study starts” and “before coming to university”); or only have university specific additions (e.g. “about the University of Twente” and “About university”). When there were combined labels, such as “Costs and tuition”, a closer look was taken at the items in the category to determine in which of these it would fit best or if it needed to stay a separate category. In the matrix, it is shown how often an item is placed in a category with a certain name. When an item is more often placed in a category this is shown in the matrix as a darker blue box. Next to the blocks, in the matrix are also vertical dark blue lines visible indicating the cluster, and light blue lines indicating the subclusters. In all cases the final choice of category name was made based on how frequently the options were used and if it covered all the different options that were included in the final label.

Ambiguities

Lastly, a look was taken at ambiguities. Ambiguities are items or groups of items found off the diagonal that have higher similarity scores and thus can cause doubt in what category they belonged. Since eyeballing darker spots (thus higher similarity scores) is known to be inaccurate (Stroes, 2018); after eyeballing, all similarity scores were checked and every score of 25% or higher was included in one of the ambiguity groups.

Results Card Sorting

To answer the first research question ("What is the average mental model of university students regarding information on the educational page of a university website?"), the average mental model of university students regarding information on the educational page of a university website is identified. To find the clusters of items of the average mental model, an item-by-item heatmap is created (Figure 2). In the heatmap, it is shown how often each of the items are found in the same category by a similarity score. When the items show a darker blue colour they have a higher similarity score, and thus were more often grouped together in the card sorting process. Items with a high similarity score are grouped together in the mental model and therefore should be put in the same category in the navigation structure of a website to increase the satisfaction and perceived usability (e.g. Dillon and Turnbull, 2005; International Standards Organisation, 1998; Nielsen, 2010; Vinney, 2021).

Here a small overview of the results will be provided, these results will be further explained in the following sections. Five main clusters are identified and can be found in Table 1. The found clusters vary in size between four to twelve items. In cluster 1 a subcluster is found and in cluster 3, three subclusters are found. This comes down to all (sub)clusters containing four or five items. The clusters are named, in order from cluster 1 to cluster 5: "Applying to university", "Finances", "About university", Student life", and "Possibilities during and after study". Next to the clusters, there are four ambiguity groups found. Ambiguity groups are groups of items that have relatively high similarity scores but are not found in the same cluster. In the following sections, firstly all clusters and subclusters will be described; following, the naming of clusters will be explained; lastly, the ambiguities are described and explained.

Description Clusters

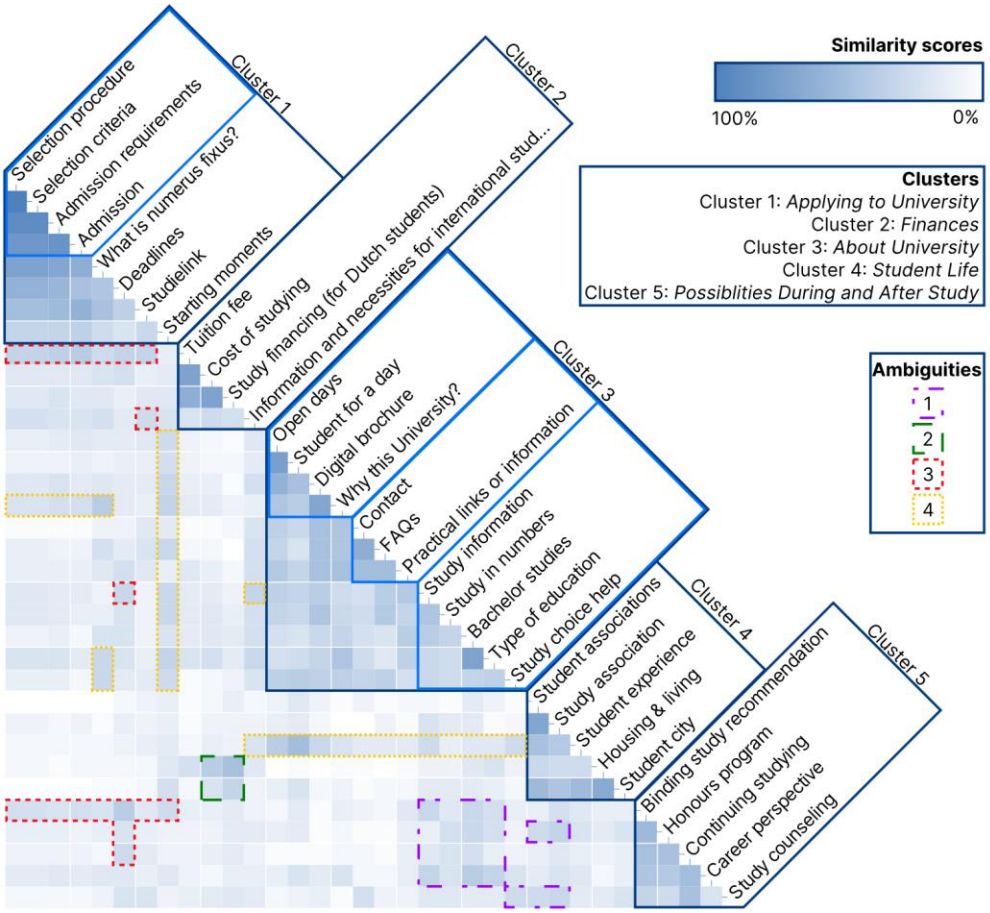
Cluster 1: Applying to university

The first cluster consists of fourteen items in total and is named "Applying to University". Within this cluster, a subcluster is found. It has been decided to put these items into a subcluster because of the high similarity scores; meaning they show a higher mutual

perceived proximity than other items in the cluster. Moreover, the items are also interpreted as fitting with each other content-wise, since they are all strongly related to enrolling in university. Lastly, when looking at the suggested names for categories there also seems to be a split where these items are more often put together in categories with certain names. Thus, because both the card sorting results, the content-wise inspection of the items, and the naming suggestions indicate a subcluster, this is created.

Figure 2

Item-by-Item Heatmap Showing the Average Mental Model of University Websites



Note. The dark blue outlines indicate main clusters. The lighter blue lines within the main clusters indicate subclusters. The other coloured and striped outlines indicate different groups of ambiguities.

The items in the subcluster are: “Selection procedure”, “Selection Criteria”, “Admission requirements”, and “Admission”. The items within this subcluster have the highest similarity scores of all (sub)clusters. The lowest being “Admission” * “Selection procedure”; and the highest “Selection procedure” * “Selection criteria” (minimum = 82%; maximum = 100%). In

the remainder of the cluster the items: “What is numerus fixus?”, “Deadlines”, “Studielink”, and “Starting moments” are included. Only one score in the cluster is seen as low (“Deadlines” * “Starting moments”). However, both items have higher similarity scores with all the other items and therefore it is decided that they should indeed be assigned to this cluster. Overall, cluster 1 has similarity scores at around 60% making this cluster conclusive (minimum =21%; maximum = 100%).

Cluster 2: Finances

When looking at the second cluster, it consists of a total of four items. Three of these items, namely “Tuition fee”, “Cost of studying”, and “Study financing” have conclusive similarity scores (minimum = 64%; maximum = 73%). However the last item “Information and necessities for international students” has low similarity scores (minimum = 21%; maximum = 24%). Therefore it seems that this item does not have much reason to be placed in this category. Therefore, Cluster 2 seems to be conclusive except for the item “Information and necessities for international students”.

Cluster 3: About university

When looking at cluster 3, this is by far the largest cluster with a total of twelve items. Since it is so large it is split up into three subclusters. The first subcluster is semantically about where prospective students can find information; the items assigned to this subcluster are: “Open days”, “Student for a day”, “Digital brochure”, and “Why this university?”. This first subcluster has higher similarity scores than the remainder of the cluster. Overall these conclusive similarity scores are around 55% (minimum = 39%; maximum = 76%).

The second subcluster contains the items “Contact”, “FAQs”, and “Practical links and information”. These are items that contain general information about the university or offer contact details where information about the university can be gathered; which matches them content-wise. The similarity scores within this subcluster are similar to the first subcluster, but a little bit lower (minimum = 45%; maximum = 64%) on average they are around 50% which is also considered a conclusive result.

The third subcluster includes five items, namely: “Study information”, “Study in numbers”, “Bachelor studies”, “Type of education”, and “Study choice help”. These items contain information about specific studies and education overall. These items have the largest variation in similarity scores (minimum = 27%; maximum = 73%). In general, these similarity scores are also lower, around 40%. Therefore it is not fully conclusive that these items should

be a subcluster. However, in the naming of clusters there also is a slight distinguishment to be seen.

Looking at cluster 3 as a whole, the similarity scores fluctuate between 18% and 76%; but are on average lower than other clusters around 35%. A cause for this large range and lower average could be the size of the cluster; when there are many items in a cluster there is always a chance that some items are placed in the same category less often. However, by dividing this cluster into three subclusters this problem is partly resolved. Moreover, the similarity scores within the cluster are still a lot higher than the similarity scores outside the cluster. Thus, it is decided this cluster should indeed be a cluster.

Table 1

Overview Clusters of Card Sorts

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
- Selection procedure - Selection criteria - Admission requirements - Admission	- Tuition fee - Cost of Studying - Study financing - Information and necessities for international students	- Open days - Student for a day - Digital brochure - Why this university?	- Student associations - Study associations - Student experience - Housing & living - Student city	- Binding study recommendation - Honours program - Continuing studying - Career perspective - Study counselling
- What is numerus fixus? - Deadlines - Studielink - Starting moments		- Contact - FAQ's - Practical links or information		
		- Study information - Study in numbers - Bachelor studies - Type of education - Study choice help		

Note. Horizontal lines between items in cluster 1 and cluster 3 indicate subclusters within clusters.

Cluster 4: Student life

Cluster 4 consists of the items “Student associations”, “Study association”, “Student experience”, “Housing & Living”, and “Student city”. Especially the items “Student associations” * “Study associations” have a high similarity score (70%); followed by “Housing & living”*“Student city” (67%). The lowest score found is 27% of “Housing & Living” * “Study associations”. The item “Housing & living” has lower similarity scores on average, but they are indecisive and not too low. Overall the items have similarity scores of around 50%; therefore, cluster 4 is considered to be conclusive.

Cluster 5: Possibilities during and after study

The last cluster is named “Possibilities during and after study” and contains the items: “Binding study recommendation”, “Honours program”, “Continuing Studying”, “Career perspective”, and “Study counselling”. This cluster seems to be the most inconclusive with similarity scores varying between 21% and 55%. However, the items in this category have stronger similarity scores with each other than with other items. Thus, cluster 5 is considered a conclusive category even though the similarity scores do not immediately indicate this.

Naming Clusters

Next to grouping the items, the participants also assigned names to the created categories. To formulate names for the clusters, a look was taken at the suggestions from participants. For the different clusters, a look was taken at the item-by-group matrix (Figure 2). Following is described how the final names for the clusters are formulated.

Based on how often a name was given to a cluster or subcluster, the names of the (sub)clusters are formulated. An overview of identified possibilities for these names is found in Table 2. When looking at the possible names for clusters, the matrix shows a relatively easy observable result for cluster 2 and cluster 4. When looking at cluster 2, two names pop out in the matrix; namely “Finances” and “Costs”. All of the items in cluster 2 are semantically involved with how to finance your study. Moreover, “Finances” was also the most suggested name for this category. Therefore cluster 2 is named “Finances”. Looking at cluster 4, there are also two category names chosen often; “Student life” and “Extra activities”. The category is named “Student life” because these items are part of living in the city where you study and the extra activities that are possible at student or study associations. Moreover, this was also the most suggested name.

When looking at the most suggested names for clusters 1, 3, and 5, there is one thing that stands out. For all three clusters, the items are often placed in the categories “Study

information" and "Information for new students". Because they are suggested for all three, it is assumed that these names do not have a distinctive enough meaning to be able to be used. For example, when students look in the category "Study information" if this was created, they could expect items from cluster 1 while there are only the items from category 5. Thus, these items are included in Table 2, but are ignored during the further process of naming.

Table 2

Suggested Names for Categories and Final Names per Cluster

Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**
"Applying to University"	*"Finances"*	*"About University"*	*"Student Life"*	*"Possibilities During and After Study"*
- Study information - Information for new students - Applying - Admission requirements	- Finances - Costs	- Study information - Information for new students - About university - General information - Study choice	- Student life - Extra activities	- Study information - Information for new students - Application process - After study - During study - Personal study life

Looking at Cluster 1 as a whole, there are three suggestions for names that were often chosen for these items; namely: "Applying", "Admission", and "Requirements". Looking at these terms, they are all different words, but have to do with the application process for university. Because there is not one suggested name that includes all the other suggestions, but they are semantically connected, the name "Applying to University" is assigned to cluster 1. Looking at the subcluster in cluster 1, the largest difference is that for these items there is a lot stronger relation found for the suggested category names. Other naming options for this subcluster could be: "Before University", "Application process", and "Selection procedure". Eventually the name "Enrolling in University". Since all suggestions are content-wise correct and it is a bit more specific than the name of the whole cluster.

Naming cluster 3 is difficult because of the large size of the cluster. Names that fit the whole cluster are: "About University", "General information", and "Study choice". Since the suggestion "About University" is most often chosen and includes the other options content-wise. Looking at the first subcluster, there are three names often indicated; "About University", the name of the whole cluster, "Information for new students", and "Study choice". "About

University" is most often assigned, however, this is not logical since it is the name of the cluster. Thus, the other two suggestions are submerged into "Information for prospective students". When naming the second subcluster, there are two names most often assigned to the items: "About University" and "General information". Since About University the name is of the whole cluster, the second subcluster is named "General information".

For the last subcluster, there are a lot of suggestions that are assigned as often as the others, although, there is one that stands out and is therefore chosen: "Study information". When naming cluster 5, there were four options found: "Application process", "After study", "During study", and "Personal study life". Looking at these options they are diverse, and there is not one option chosen significantly more than the others. Therefore, a look is taken at the meaning of the items in combination with the possible assigned names. While doing this, especially "Application process" seems to be out of place. All the items have to do with information that is usually used during or after your studies. The items do seem to match in meaning with the other suggestions "After study", "During study", and "Personal study life". Because no suggestion stands out as the obvious choice, these names are combined into the final name: "Possibilities during and after study".

Ambiguity Groups

Ambiguity groups are groups of items not placed on the diagonal axes but still have similarity scores of 25% or higher. The ambiguities are identified based on their colour that is darker than the adjacent blocks, and after that checked box-by-box since only identifying by colours can be inaccurate (Stroes, 2018). Ambiguities that are placed in the same group have the same coloured border (see Figure 2). The following paragraphs will describe the ambiguity groups in more detail.

Ambiguity 1: Study program

The first found ambiguity group is considered to have items that are all related to the study program. When looking at the items it mainly is a block of items from the subcluster "Study information" and the cluster "Possibilities during and after study". These are both related to the study program. Moreover, also the items "Student association" and "Study association" are seen as ambiguous with items from the cluster "Possibilities during and after study". Looking at these items, they are not only content-wise related to the study program, but also grammar-wise related to each other. All mentioned items have either "student" or "study" in them. This is interesting to see since there is a large number of people grouping the items in

the clusters they are now, but there are also people who expect the items placed in the syntactically related place.

Ambiguity 2: Financial

The second ambiguity group contains three items “Study financing”, “Information and necessities for international students”, “Housing & living”, and “Student city” which all have high similarity scores with each other. This ambiguity group is considered to have the main factor of financials as a content-wise cause for the ambiguity. This is because these items are all related to either the costs of living in a student city and how to finance this.

Ambiguity 3: Formal obligations for university

All ambiguities for the third group are connected to the cluster “Applying to university”. The items “Tuition fee” and “Binding study recommendation” show similarity scores over 25% with (almost) the whole cluster which indicates that at least a quarter of the participants grouped these items with the items from the cluster “Applying to university”. Other items that show inconclusive similarity scores are: “Information and necessities for international students”, “Study information”, “Honours program”, and “Continuing studying”. Looking at these items, there does not seem to be a semantic connection, therefore, to explain the ambiguity there must be a cause in content or interpretation. What stands out when looking at the similarities, most of the individual items (not with the whole cluster) have inconclusive similarity scores with the item “Deadlines”. This could either indicate that people doubt where the item “Deadlines” should be placed or that the item “Deadlines” is unclear in what it means. That “Deadlines” is unclear is a likely scenario, since this was also indicated in the comments section of the card sorting task. Therefore, the items with similarity scores with “Deadlines” might be vague why they are ambiguities and the problem could lie in the interpretation of the item.

Since some of the items are vague, the emphasis while analysing the content of the items will be on the combination of cluster one “Applying to university”, and the items “Tuition fee”, and “Binding study recommendation”. These items all have to do with formalities that have to be met for university. The first cluster is applying to university, which is considered a formal process, therefore it is logical to connect other requirements to this; “Tuition fee” and “Binding study recommendation” are necessary processes to be able to keep studying after applying. Therefore, there might be a content-wise similarity why they are often put in the same category.

Ambiguity 4: Study and university choice

In the last ambiguity group, all items are connected to the cluster “About university”. Items that stand out are “Starting moments” and “Student experience”, these both have

inconclusive similarity scores with all items from the cluster “About university”. Other ambiguities are found with the items “What is numerus fixus?” and “Information and necessities for international students”. Another notable ambiguity is that of the item “Why this university?” with the cluster “About university”; there are inconclusive similarity scores with many items from this cluster. It is chosen to put this ambiguity in this group and not the ambiguity group “Formal obligations for university” because there does not seem to be a content-wise connection between the item and the remainder of the ambiguity group.

While looking at why this ambiguity group exists, there is again no semantic reason found. However, considering the meaning of these items, they are all considered to have an influence on either study or university choice. For example, when the starting moments are and what the student experiences are at this university.

Methods Perceived Usability Test

Participants

In this study eleven participants took part. The participants were a convenience sample of acquaintances of the researcher; they were recruited via WhatsApp or phone. The perceived usability test took place at a set moment and place on the researcher’s laptop. Five of the participants identified as male and six as female. All participants were Dutch. Three participants were aged 18-21, seven were 22-25, one was older than 25.

Materials

For this study, multiple materials were used. The first materials were the university websites on which tasks were performed: tue.nl/en/education and tudelft.nl/en/education. The choice was made to use the other two technical universities in The Netherlands next to the University of Twente (UT) because, firstly, it was assumed these websites would not be in daily use by the participants. Secondly, the content would be most similar; instead of e.g. a technical and a socially focussed university. Furthermore, two sets of tasks for the participants were created based on being compatible with the mental model (see ‘Results Card Sorting’) or not being compatible with the mental model. The tasks were based on items that had relatively high similarity scores, and thus were clear in which cluster they should fit. The tasks on the website from Delft University of Technology (TU Delft) were compatible with the mental model, thus the information was placed on the website where people expect it to be. The tasks were:

- *“You are thinking about doing your masters in architecture, Urbanism, and Building sciences; what are the admission requirements?”*

- *"You are going to do your masters in Delft next September; what is your tuition fee going to be?"*
- *"What are fun activities to do in Delft as a student city?"*

The information for the tasks on the website of Eindhoven University of Technology (TU/e) was placed at another place than the mental model indicated. The tasks for this website were:

- *"You are not happy with your current study and want to start a new bachelor's in computer science next year. How does the selection procedure work?"*
- *"What is a proper estimation of your living expenses in Eindhoven, excluding tuition fee?"*
- *"Where can you find more information about Eindhoven as a student city?"*

The provided tasks were performed on the laptop of the researcher. The tasks were timed by the researcher with a stopwatch on their phone. The time spend on the task, and whether or not the task was completed was noted by the researcher.

When the tasks on the first website were completed the participant received the System Usability Scale (SUS) in the program 'Qualtrics' (Appendix J). It consisted of ten statements that the participant has to reply to on a Likert-like scale from "Strongly disagree" to "Strongly agree". The SUS is a questionnaire that measures the perceived usability of a web application (Brooke, 2013). The SUS had first been used in 1996 and has proved to be a reliable and valid instrument to measure perceived usability (Lewis, 2017). Moreover, the SUS is also usable as a valid instrument in small samples and a quick method to measure the perceived usability (Brooke, 2013).

Procedure

When starting, the participant was first provided with the Informed Consent (Appendix I) and was explained that they were going to take part in a perceived usability test, perform tasks on websites, and answer some questions about that. It was made clear to the participant that they could quit at any moment and the website was being assessed and not their abilities. After consenting to take part in the study the participant answered some demographic questions about their gender, nationality, and age.

The participant took place at a desk with the researcher's laptop on the educational page of the first university website in front of them. Half of the participants started at the website of the TU Delft and the other half started at the website of TU/e so that order would not have an effect. The participants were instructed to perform the task as they would normally look for

information. However, they were asked to only navigate through the menus and not use the search bar. Moreover, the participants could give comments to the researcher about the task or website during their task if they wanted to.

The researcher read the first task aloud, and when confirmed the participant understood the task, they started to try to perform the task. During the task, the researcher timed how long it took to perform the task and whether it was completed. If a participant took longer than five minutes to complete a task, the task was stopped. When finished, the researcher checked if the participant was ready for the next task or needed a break; if they were fine the next task was provided, this took place after every task. When all three tasks of the first website were finished, the researcher presented the Qualtrics questionnaire with the SUS to the participants to fill out. The process of the first website was repeated at the second university website. After filling in the SUS for the second time the participant was asked which website they liked best. After that, they were finished and thanked for their participation.

Data Analysis

Starting the data analysis, the dataset was inserted into R and a demographic analysis on age, gender, and nationality took place. To perform the analysis of the data collected from the SUS, the data was imported into the System Usability Scale Analysis Toolkit. The toolkit developed by Blattgerste et al. (2022) was developed to support in the analysis of the SUS and is an open source web-based for single and multivariable studies. Since two results from different websites were analysed and compared, this is a multivariable study. In this program, the standardized scores are immediately calculated, compared to metadata, and shown in graphs. Moreover, based on the study of Sauro and Lewis (2016) an industry benchmark is provided for different scores to place the results in context for usability in the industry. Next to the benchmark, the percentile score is provided that tells how the SUS-score is compared to other systems evaluated with the SUS. Lastly, a conclusiveness score is provided (Stetson & Tullis, 2004). This score shows how conclusive the SUS results are with the current sample size, thus if the chance is large that the score would change with a larger sample (Stetson & Tullis, 2004).

Results Perceived Usability Test

To answer the second research question (*"To what extent does the perceived usability of a website increase for university students if the website matches the average mental model of a student?"*), a comparison was made between the perceived usability of a website where participants performed tasks according to the mental model and a website where participants

performed tasks that did not line up with the mental model. The perceived usability was measured using the System Usability Scale (SUS).

Table 3

Results of SUS Analysis Showing Mean SUS Score, Standard Deviation (SD), Minimum and Maximum Score, and Conclusiveness of the Results

Variable	*M*	*SD*	Minimum score	Maximum score	Conclusiveness
TU Delft	66.36	14.43	37.50	87.50	98%
TU/e	71.59	6.15	60.00	82.50	98%

There was a mean standardized SUS-score calculated for the TU Delft (*M* = 66.36, *SD* = 14.43) and the TU/e (*M* = 71.59, *SD* = 6.15). Looking at the boxplot (Figure 3), the scores of the TU Delft vary more (Minimum = 37.50; Maximum = 87.50) than the score of the TU/e (Minimum = 60.00; Maximum = 82.50). Of the participants, six indicated liking the website of the TU Delft more, four indicated liking the website of the TU/e better, and one did not have a preference. All participants completed all six tasks within the five minute time-frame.

Figure 3

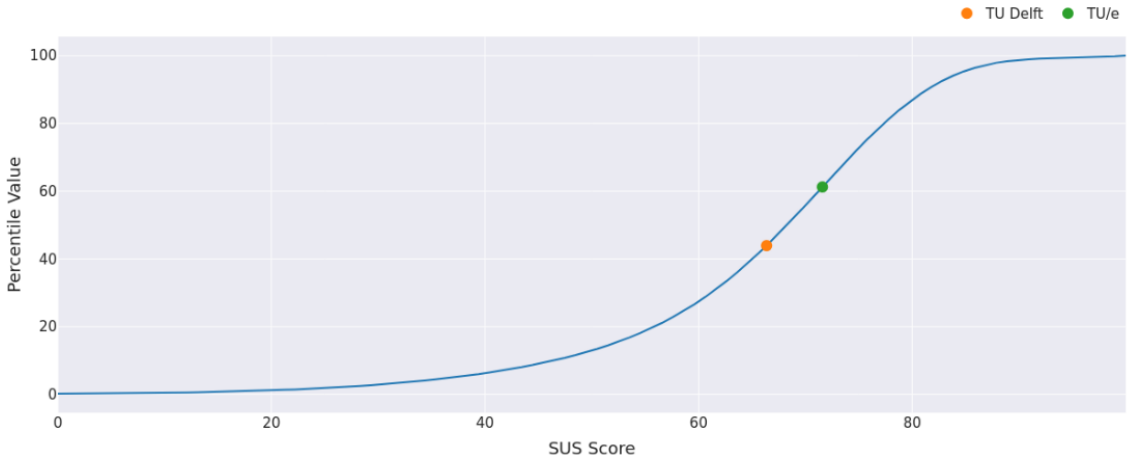
Boxplots of the System Usability Scale Score from the Websites of the TU Delft and TU/e

Note. The boxplots show the spread of the SUS scores for the websites of the TU Delft and TU/e.

The SUS-scores results in benchmarks of; for the TU Delft *"below average"*, for the TU/e *"Above average"*. Putting the scores in comparison to other SUS-scores from meta show percentile score of 43.94% for the TU Delft and 61.26% for the TU/e (see Figure 4). The results of this are considered conclusive, since a conclusiveness score of 98% is found.

Figure 4

Percentile Values of the System Usability Scale Scores of the TU Delft and TU/e.



Discussion

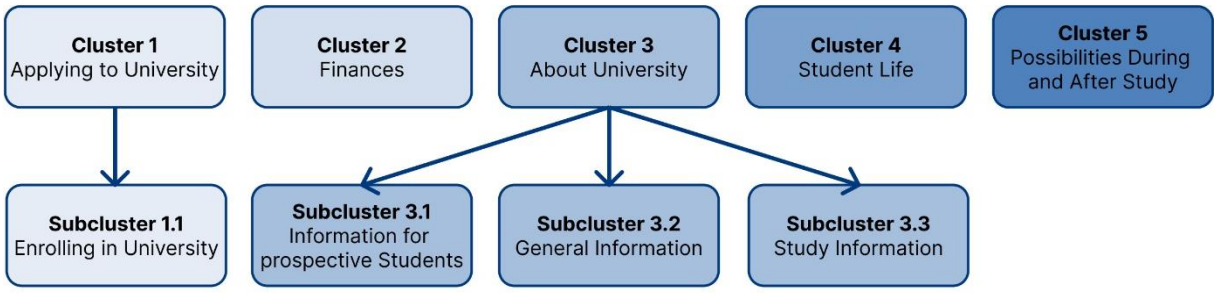
The purpose of this research is to elicit the mental model of university students for the educational part of university websites and to determine whether the perceived usability would increase if the information architecture of a website is similar to this mental model. This aim is reached by carrying out a card sorting study and a perceived usability test. In the following sections the results of both these studies will be discussed on a higher level and in the light of literature; furthermore, the limitations of the research will be identified, recommendations for future research are made and a conclusion is formulated.

Findings

The first research question, *"What is the average mental model of university students regarding information on the educational page of a university website?"*, is answered by the created cluster structure based on the heatmap from the card sorting task. In Figure 5 an overview is shown of the proposed structure with five clusters and four subclusters with their corresponding names. The items that should be placed in each of the clusters can be found in Table 1. When looking at the information architecture of the educational page of a university website, this should ideally be designed based on the structure described in Figure 5 The structure is based on the mental model, which means that university students will, on average, expect information to be in this place. Putting information in a place where people expect it, is crucial to have an easy-to-use website (Sinha & Boutelle, 2004). Sinha and Boutelle (2004) explain that the information architecture should be user-centred and therefore be based on the user's expectations, thus the mental model.

Figure 5

Graphic Overview of the Proposed Information Architecture for the Educational Page of University Websites.



Note. Content of clusters and subclusters is defined in "Results Card Sorting".

Based on this cluster structure a perceived usability test is performed. The aim of performing the perceived usability test is to determine whether the perceived usability is higher when tasks are performed that are according to the mental model than when they are not congruent. This also answers the second research question; *"To what extent does the perceived usability of a website increase for university students if the website matches the average mental model of a student?".* Two sets of tasks are compared, and the perceived usability is measured using the SUS. The results show that on average the website of the TU/e scored better than the website of the TU Delft. This is contrary to the expectation that a website that matches the average mental model would have a higher perceived usability. However, it is noticeable that when participants indicated which website they like better through a direct question the majority pointed out the website of the TU Delft; opposed to the results of the SUS.

The results of the SUS being opposite to the question of preference means that no indication is found that the mental model has an effect on the perceived usability while browsing through a university website. This is in line with the research of Schmettow and Sommer (2016) who also did not find a relation between the measures of the mental model and the usability.

Design Recommendations

Looking at the elicited mental model in comparison to the information found on university websites some things stand out. The found structure consists of five groups that would be the main menu. However, most existing university websites include more items in their menu (e.g. TU/e has seven menu items, the UT has eleven, RU has six, MU also has eleven menu items). This indicates that existing websites have a broader menu structure and the end-

users would prefer a deeper menu structure. Moreover, Juncan (2013) investigated for people of different ethnicities if they prefer a broad or deep information structure. She found that people with a western background prefer and have a better browsing performance on a website with a deep information structure. Thus, it could be supported that for a website of a Dutch university, where mostly students with a western background study, a deep information structure should be used. This means that the found mental model with less items would be a better basis for the information architecture.

Considering the existing menus, another aspect stands out, namely the heading that consists of items that provide more details about the education being taught at that university. This item is found on almost every university website (from Appendix A). However, this cluster is not found while looking at the mental model. In the mental model these items are mostly spread around cluster 1 (Applying to university), cluster 2 (Finances), and subcluster 3.3 (Study information). Since in the mental model there was no cluster found with this name, it is assumed that is better to not have this category anymore.

A second thing that is noticed when looking at the existing website is that most of the websites have at the beginning of the educational page a split between bachelor and master; sometimes also pre-masters and other educational programs are included. During the tasks of the perceived usability test it is observed that the participants all use the split between bachelor and master almost directly and efficiently when looking for information about a certain study programme. This observation supports the already existing structure of the university websites in Appendix A. Therefore, a recommendation would be to keep part of the information architecture even though it is not found in the mental model.

Another remarkable aspect of the menus is the category “Student life”. This cluster was identified by participants to be a separate cluster. However, none of the university websites that are used as a basis for the cards (see Appendix A) have a menu item called “Student life” or a category with a similar name that includes items from this cluster. Therefore, a recommendation would be that there should be a menu item in the range of student life with the items of the cluster.

While looking at the ambiguities of the mental model, there is one ambiguity that stands out most. Namely the ambiguity between subcluster 3.3 (Study information) and cluster 5 (Possibilities during and after study). Since there is such a large block of items that are considered ambiguous it could be argued that on a website there should be a crosslink between subcluster 3 and cluster 5. Therefore, a design recommendation is to offer a connection between

these two clusters since there is a relatively large chance people will look at the wrong spot for information from this ambiguity.

During the perceived usability test, the participants indicated that the website of the TU Delft contains too much text. When asking questions about this participants explain that the webpages itself contain a lot of information and long pieces of text; this is difficult to scan through while looking for certain information. Furthermore, the general menu on the educational page is described as too large; it is hard to gain an overview of the menu.

On the website of the TU/e another problem is noticed; namely the menu items itself. The items are described by the participant as confusing and unclear. Multiple times it is mentioned that information could be in at least three headings. The heading are considered too vague and to find out what information they contain you have to be hovering over the heading and see what items appear. This finding that there are too many menu items is also supported by the theory of Juncan (2013) that people with a western background prefer a deeper information structure over a broad information structure. Thus, the advice is to have less and clearer menu items that are more distinguished.

Limitations

In the first part of this research, the card sorting, several limitations are found that need to be discussed. Firstly, during the cleaning of the dataset, a total of twelve participants is removed from the sample due to having vague categories or having only one or two categories. These participants are removed to have better, less ambiguous results in the heatmap. However, there could also be a reason that people only created two groups. It is chosen to not include these people, because for a cluster structure of a website more categories are desired. Not including the information of participants who only created two categories provides a heatmap with less ambiguous results. Therefore, it is decided to not include these participants in the final sample. More problematic could be the exclusion of participants with vague category names. These participants were only removed if the items in the vague categories are not comparable to results of other participants, as judged by the researcher. However, the selection criteria are probably too strong since there is a total of twelve participants excluded out of 45; which is a lot. Due to the constraints of time it is decided to not try an analysis with the participants with vague categories. However, it would be recommended in future research that participants with vague categories are included for the determination of the cluster structure.

Secondly, the understanding of the cards could also be problematic. Some of the cards were perceived as ambiguous or unclear by the participants. For example, the item “Deadlines”

was mentioned several times in the comments section as people did not know how to handle this item. It was chosen specifically to not have explanations on the cards, as there would also not be an explanation on the website when navigating through this. However, some items also gained clarity from the location at the website; when looking at deadlines this is placed in the menu item of application process. To avoid this, in a new set of items the items that are vague could be made clearer by changing them; "Deadlines" could for example become "Application Deadlines". Another option would be to do include descriptions of the items even though this is not shown on a website. Since this did both not happen, this is seen as another limitation.

In the second part of this research, the perceived usability test, also several limitations are identified. Firstly, there is the limitation that there are only two websites used. Thus, this could mean that the results are dependent on other factors of the website and not necessarily the information architecture that is, or is not congruent with the mental model. Moreover, since there are multiple tasks performed on the same website, learning effects will likely occur (Schmettow & Sommer, 2016). Ideally, the tasks would have been performed on more websites, and every website would have been tested with tasks that are both in line and not in line with the mental model by different participants. Another option is to let the participants perform only one task per website on more different websites as was done in the experiment of Schmettow and Sommer (2016) to avoid the learning effects. However, due to time constraints, these options are not executed.

Moreover, a problem could occur with the formulation of tasks. The tasks are now formulated based on the location on the website that is corresponding with the mental model or not. However, there is still a difference in the extent to which an item is in the right place according to the mental model. The place of an item with a higher similarity score is more certain than the place of an item with a lower similarity score. Thus the usage of a gradient that takes the height of the score into account would be better (e.g. the mismatch score as used by Schmettow and Sommer (2016)).

Another limitation that could be argued in the card sorting study is the number of cards. Where some research suggests 30-40 cards (Card sorting, 2013), others suggest that 40-80 are necessary to have a complete representation of a website (Sherwin, 2018). Thus, it could be that with a larger sample of cards other clusters would have been created, therefore, it could be a limitation. This problem could be solved by performing the study again with a larger sample of item cards.

When looking more closely at the university websites in general, most of the websites have at the beginning of the educational page a split between bachelor and master; sometimes

also pre-masters and other educational programs are included. However, having the items on the cards that were now chosen, this difference between master's and bachelor's studies was difficult to make. It is a limitation to this study that this distinguishment is difficult to make with these items, since it is not known whether this is also found in the mental model of university students. By choosing more or other items on the cards, this distinguishment could have occurred. Another way to solve this limitation would be including a hierarchical card sort. In this type of card sorting, participants sort the card on multiple levels, as often can be found in an information architecture (Davies, 2008). When doing this, items like "Bachelor" or "Master" can be included, or other items that could make sense to distinguish in. Another option would be to use a closed card sort and include the categories "Bachelor" and "Master". There could also be a combination of a closed hierarchical card sort.

The constraints of time also lead up to the next limitation, namely the representativeness of the sample, in both studies. The sample for these researches consists of acquaintances of the and psychology students via Sona Systems, which makes it not representative of the population of university students in the Netherlands. Looking at the background of the participants, people who are students in different studies might have a different mental model but since Sona Systems is used most of the students are psychology students. Also, nationality is not generalizable to the whole population; in the card sorting study most of the participants were Dutch or German. In the perceived usability test, all participants are Dutch. In The Netherlands there are many more nationalities found within the student population. Lastly, the sample for the perceived usability test is also really small. Even though the SUS gives valid results for samples of this size, it could be argued that a larger sample would have given better results.

Future Research

After this research, some recommendations are made for possibilities in future research projects. To address first, looking at the results of the perceived usability test, there does not seem to be a connection between the mental model and the perceived usability. Thus, it could be argued that before doing more research on the topic of the determining the mental model it should be validated that there is a connection; otherwise the results are not usable. Counting on the fact that this effect can be validated, suggestions are also made for future card sort studies.

Firstly, the cards sorting research could be repeated with a more representative sample in cards as well as participants. When the research is performed with a more wide variety of students, results are expected to be more representative for the whole of the student population. Moreover, when repeating the card sorting, a closer look should be taken at the items on the

cards. Some items might need a little clarification; for example, with the item “Deadlines” it should be included which deadlines are meant to get more informative results. Also more items could be added to see if other clusters are created and represent more of the website as Sherwin (2018) suggests.

If the card sort study would be repeated, the same or with the before mentioned adjustments, it would be recommended to do either a closed card sort or a hierarchical card sort. By performing one of these other forms, the distinguishment between bachelor and master tracks can be made. This way it can be checked if and how this takes place since this was found in the mental model, as this is observed in the existing websites as well as during the perceived usability test.

While looking at the follow-up, the perceived usability test, there are more possibilities to determine the (perceived) usability. A first recommendation would be to include more websites and place tasks and to switch between being congruent with the mental model and not. By doing this, the effect of a website can have is taken away and only the information architecture is compared. To do this more rigorously, a website could be built according to the mental model. This way the whole website will be congruent with the mental model and can then be compared to other (existing) websites.

Another option would be to perform different usability tests. As the SUS only measures the perceived usability and not the actual usability this is also an interesting aspect to look at. For example, the path to an information goal or the number of clicks could indicate the usability of a website.

Conclusion

In this study, firstly the mental model of university students regarding the educational page of university websites is elicited and it is determined that integrating this into the information architecture has no effect on the perceived usability. Having found this mental model, can add new knowledge to future designs of university websites, since this is not yet found when searching for literature. Where most items have similarity scores that place them in clusters clearly, there is doubt about where some of the more vague items belong and some ambiguities are found. However, since the perceived usability test did not show a relation between the mental model and the perceived usability, it should first be validated that building an information architecture based on the mental model has positive effects. Nevertheless, based on the findings from both researches recommendations for the information architecture are

made. Moreover, suggestions are made that can be integrated into future research adding more knowledge to this field of research.

References

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction*, *24*(6), 574-594. https://doi.org/10.1080/10447310802205776

Blattgerste, J., Behrends, J., & Pfeiffer, T. (2022). *A web-based analysis toolkit for the system usability scale*. https://sus.mixality.de/

Branaghan, R. J., Covas-Smith, C. M., Jackson, K. D., & Eidman, C. (2011). Using knowledge structures to redesign an instructor–operator station. *Applied ergonomics*, *42*(6), 934-940. https://doi.org/10.1016/j.apergo.2011.03.002

Brooke, J. (2013). SUS: a retrospective. *Journal of usability studies*, *8*(2), 29-40. https://doi.org/10.5555/2817912.2817913

Budio, R. (2017). *Quantitative vs. qualitative usability testing*. Nielson Norman Group. https://www.nngroup.com/articles/quant-vs-qual/

Bussolon, S., Russi, B., & Missier, F. D. (2006). Online card sorting: as good as the paper version. Proceedings of the 13th Eurpoean conference on Cognitive ergonomics: trust and control in complex socio-technical systems, 113-114. https://doi.org/10.1145/1274892.1274912

Bussolon, S. (2009). Card sorting, category validity, and contextual navigation. *Journal of Information Architecture*, 1(2) 5-29.

Card Sorting. (2013). Department of Health and Human Services. https://www.usability.gov/how-to-and-tools/methods/card-sorting.html#:~:text=Best%20Practices%20for%20Card%20Sorts,especially%20for%20a

Case, D. O., & Given, L. M. (2016). Looking for information: A survey of research on information seeking, needs, and behavior.

Davies, R. (2008). *Hierarchical card sorting*. Monitoring and Evaluation News. https://mande.co.uk/special-issues/hierarchical-card-sorting-hcs/

Diefenbach, S., Kolb, N., & Hassenzahl, M. (2014, June). The'hedonic'in human-computer interaction: history, contributions, and future research directions. In *Proceedings of the*

2014 conference on Designing interactive systems (pp. 305-314). https://doi.org/10.1145/2598510.2598549

Dillon, A., & Turnbull, D. (2005). Information architecture. In McDonald, J. D., & Levine-Clark, M. (4th ed.) CRC Press: Boca Raton. https://doi.org/10.1081/E-ELIS4

Faiks, A., & Hyland, N. (2000). Gaining user insight: a case study illustrating the card sort technique. *College & research libraries*, *61*(4), 349-357. https://doi.org/10.5860/crl.61.4.349

Flavián, C., Guinalíu, M., & Gurrea, R. (2006). The role played by perceived usability, Satisfaction, and consumer trust on website loyalty. *Information & management*, *43*(1), 1-14. https://doi.org/10.1016/j.im.2005.01.002

Hudson, W. (2005). Playing Your Cards Right. *Interactions* 12 (5): 56–58. https://doi.org.com/10.1145/1082369.1082410.

ISO. (1998). Ergonomic requirements for office work with visual display terminals (vdts)–part 11: Guidance on usability. ISO Standard 9241-11: 1998 (Vol. 55).

Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences*, *107*(43), 18243-18250. https://doi.org/10.1073/pnas.1012933107

Juncan, L. (2013). Website Globalization Strategy: A Cross-Cultural Analysis of Website Structure.

Lah, U., Lewis, J. R., & Šumak, B. (2020). Perceived usability and the modified technology acceptance model. *International Journal of Human–Computer Interaction*, *36*(13), 1216-1230. https://doi.org/10.1080/10447318.2020.1727262

Lewis, J. R. (2018). The system usability scale: past, present, and future. *International Journal of Human–Computer Interaction*, *34*(7), 577-590. https://doi.org/10.1080/10447318.2018.1455307

Lewis, J. R., Utesch, B. S., & Maher, D. E. (2015). Measuring perceived usability: The SUS, UMUX-LITE, and AltUsability. *International Journal of Human-Computer Interaction*, *31*(8), 496-505. https://doi.org/10.1080/10447318.2015.1064654

Madejska, A., & Persson, A. (2019). *Menu anchor interactors in mobile website contexts: The*

perceived usability of menu navigation on three different types of websites. https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1330799&dswid=-1309

Morville, P., & Rosenfeld, L. (2006). *Information architecture for the World Wide Web: Designing large-scale web sites*. " O'Reilly Media, Inc.". https://books.google.nl/books?hl=en&lr=&id=2d2Ry2hZc2MC&oi=fnd&pg=PR5&dq=information+architecture&ots=oohexl_PAe&sig=qN4YbQnmDou3ZojpH_gI0UetkHw#v=onepage&q=information%20architecture&f=false

Nakhimovsky, Y., Schusteritsch, R., & Rodden, K. (2006, April). Scaling the card sort method to over 500 items: restructuring the Google AdWords Help Center. In *CHI'06 Extended Abstracts on Human Factors in Computing Systems* (pp. 183-188). https://doi.org/10.1145/1125451.1125491

Nielsen, J. (2010). *Mental Models*. https://www.nngroup.com/articles/mental-models/

Otter, M., & Johnson, H. (2000). Lost in hyperspace: metrics and mental models. *Interacting with computers*, *13*(1), 1-40. https://doi.org/10.1016/S0953-5438(00)00030-8

Paul, C. L. (2008). A modified delphi approach to a new card sorting methodology. *Journal of Usability studies*, *4*(1), 7-30.

Righi, C., James, J., Beasley, M., Day, D. L., Fox, J. E., Gieber, J., . . . Ruby, L. (2013). Card sort analysis best practices. 8(3), 69-89.

Rook, L. (2013). Mental models: A robust definition. The Learning Organization, 20(1), 38–47. https://doi.org/10.1108/09696471311288519

Roshan, H., & Ahmadi, P. A (2022) Practical Model to Increase Usability of Website Navigation by User Experience Design Strategies. *SSRN*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4219646

Schmettow, M., & Sommer, J. (2016). Linking card sorting to browsing performance – are congruent municipal websites more efficient to use? Behaviour & Information Technology, 35(6), 452-470. https://doi.org/10.1080/0144929x.2016.1157207

Sauro, J., & Lewis, J. R. (2016). *Quantifying the user experience: Practical statistics for user research*. Morgan Kaufmann. https://doi.org/10.1145/2413038.2413056

Schall, A. (2014). Information architecture and web navigation. In Eye tracking in user

experience design (pp. 139-162). Elsevier. https://doi.org/10.1016/B978-0-12-408138-3.00006-6

Sherwin, K. (2018). *Card Sorting: Uncover Users' Mental Models for Better Information Architecture*. Nielsen Norman Group. https://www.nngroup.com/articles/card-sortingdefinition/

Sinha, R., & Boutelle, J. (2004). Rapid information architecture prototyping. ACM Press. https://doi.org/10.1145/1013115.1013177

Stetson, J. N., & Tullis, T. S. (2004). A comparison of questionnaires for assessing website usability. *UPA Presentation*.

Stroes, J. D. (2018). Adjusting the navigation structure of a faculty intranet to the average user with Card Sort and Q-sort. University of Twente.

Tullis, T., & Wood, L. (2004, June). How many users are enough for a card-sorting study. In *Proceedings UPA* (Vol. 2004). Minneapolis (EUA): Usability Professionals Association (UPA).

Vinney, C. (2021, July 21). *What are mental models and how are they used in UX Design?* CareerFoundry. https://careerfoundry.com/en/blog/ux-design/mental-models-ux-design/

Westbrook, L. (2006). Mental models: a theoretical overview and preliminary study. Journal of Information Science, 32(6), 563–579. https://doi.org/10.1177/0165551506068134

Wood, J. R., & L. E. Wood. (2008). Card Sorting: Current Practices and Beyond. *Journal of Usability Studies* 4 (1): 1–6.

Appendices

Appendix A – Selected Dutch Universities with Corresponding Websites

University	Abbreviation	Website
University of Twente	UT	https://www.utwente.nl/
Rijksuniversiteit Groningen	RUG	https://www.rug.nl/
Radbouduniversiteit Nijmegen	RU	https://www.ru.nl/
Eindhoven University of Technology	TUE	https://www.tue.nl/en/
Universiteit van Amsterdan	UvA	https://www.uva.nl/
Leiden University	LU	https://www.universiteitleiden.nl/en
Delft University of Technology	TUD	https://www.tudelft.nl/
Maastricht University	UM	https://www.maastrichtuniversity.nl/

Appendix B – Found Possible Items on University Sites

Binnen bachelors algemeen	**UT**	**RUG**	**RU**	**TUE**	**UvA**	**LU**	**TUD**	**UM**
Open dagen	X	X	X	X	X	X	X	X
Bacheloropleidingen	X	X	X	X	X	X	X	X
Studenten ervaringen	X	-	X	X	X	X	X	X
Digitale brochure	X	X	-	X	-	X	X	X
Waarom deze uni	X	X	X	-	X	X	-	X
Practische links / info	X	-	-	-	X	-	X	X
Instroommomenten	X	X	X	-	X	X	X	X
Type onderwijs	X	X	-	X	-	X	-	X
Beoordeling en toetsen	X	-	-	-	-	-	-	-
Studiebegeleiding	X	X	X	X	X	X	X	X
Keuzevrijheid	X	-	-	X	X	-	-	-
Studievoorlichting	X	-	X	X	X	X	X	-
Studie in cijfers	X	-	X	X	-	-	X	X
Hulp bij studiekeuze	X	X	X	-	X	X	X	X
Tips en adviezen	X	-	-	-	-	X	-	X
Inschrijven	X	X	X	X	X	X	X	X
Selectieprocedure	X	X	X	X	X	X	X	X
Selectiecriteria	X	X	X	X	X	X	X	X
Toelatingseisen	X	X	X	X	X	X	X	X
Bindend studieadvies	X	X	X	X	X	X	X	X
Collegegeld	X	X	X	X	X	X	X	X
Deadlines	X	X	X	X	X	X	X	X
Doorstuderen	X	X	X	X	X	X	X	X
Kansen op de arbeidsmarkt	X	X	X	-	X	X	X	X
De studentenstad	X	X	-	-	X	X	X	X
Studentenvereniging	X	X	X	-	X	X	X	-
Contact	X	X	X	X	X	X	X	X
Honoursprogramma	X	X	X	X	X	X	X	X
Meeloopdag	X	X	X	X	X	X	X	X
Studielink	X	X	X	X	X	X	X	X
FAQ’s	X	X	-	X	X	X	-	-
Wonen en leven	X	X	X	X	X	X	X	X
Wat kost studeren?	X	-	X	X	X	-	X	X
Benodigdheden voor internationale student (Registratie in de stad, bankrekening, visa, verzekering, etc.)	X	X	-	X	-	-	X	X
Studieverenigingen	X	X	-	X	X	-	X	X
Studiefinanciering	X	-	X	-	X	-	X	X
Numerus fixus	X	X	X	-	X	X	X	X
Scholarships	-	-	-	X	X	-	X	X
Topsport en studeren	-	X	-	-	X	-	X	-

Appendix C – Final Item Selection for Cards

Present on all websites	
Open dagen	Open days
Bacheloropleidingen	Bachelor studies
Studiebegeleiding	Study counselling
Inschrijven	Admission
Toelatingseisen	Admission requirements
Selectieprocedure	Selection procedure
Selectiecriteria	Selection criteria
Bindend studieadvies	Binding study recommendation
Collegegeld	Tuition fee
Deadlines	Deadlines
Doorstuderen	Continuing studying
Contact	Contact
Honoursprogramma	Honours program
Meeloopdag	Student for a day
Studielink	Studielink
Wonen en leven	Housing & living
Present on 7 out of 8 websites	
Studenten ervaringen	Student experience
Instroommomenten	Starting moments
Hulp bij studiekeuze	Study choice help
Kansen op de arbeidsmarkt	Career perspective
Hoe werkt numerus fixus	What is numerus fixus?
Present on 6 out of 8 websites	
Digitale brochure	Digital brochure
Waarom deze uni	Why this University?
Studievoorlichting	Study information
De studentenstad	Student city
Studentenverenigingen	Student associations
Studievereniging	Study association
Wat kost studeren	Cost of studying
Present on 5 out of 8 websites	
Practische links / info	Practical links or information
Type onderwijs	Type of education
Studie in cijfers	Study in numbers
FAQ’s	FAQs
Benodigdheden voor een internationale student	Information and necessities for international students
Studiefinanciering	Study financing (for Dutch students)

Appendix D – Informed Consent Card Sorting

Welcome to our study!

Please read this page carefully for if you choose to proceed, you declare to understand and consent to the described information.

You are being invited to participate in a research study titled "UNTANGLING THE MIND increasing website search efficiency through card sorting". This study is being done by Sander Overkamp, Linda Renes, and Laura Scharstuhl from the Faculty of Behavioural, Management, and Social Sciences at the University of Twente.

Goal of the study
The purpose of this research study is to attain the mental models of students regarding navigating university websites and will take you approximately 30 minutes to complete.

The experiment
You will participate in a study where we will gather information using a card-sorting experiment. More instructions on how the experiment works will follow on the next page.

Benefits and risks of participating
There are no physical, legal or economic risks associated with your participation in this study. This experiment has been reviewed and approved by the BMS Ethics Committee/domain Humanities & Social Sciences. However, as with any online-related activity, the risk of a breach is always possible. To the best of our ability, your answers in this study will remain confidential.

Voluntary participation and withdrawal
Participation in this study is completely voluntary. As a participant, you may stop your participation in the study at any time or refuse to allow your data to be used for the study, without giving any reasons. Stopping participation will not adversely affect you or any compensation already received. If you decide to discontinue participation during the study, the data you have already provided will be used in the study until consent is withdrawn.

Do you want to stop the study, or do you have questions and/or complaints? Please contact one of the researchers:

- Sander Overkamp: s.overkamp@student.utwente.nl
- Linda Renes: l.e.m.renes@student.utwente.nl
- Laura Scharstuhl: l.w.scharstuhl@student.utwente.nl

Data Confidentiality
We make every effort to protect your privacy to the best of our ability. No confidential information or personal data from or about you will be disclosed in any way that will allow anyone to recognize you.

Before our research data is released to the public, your data will be anonymized as much as possible, unless you have given explicit permission in our consent form for your name to be mentioned, for example in a quote.

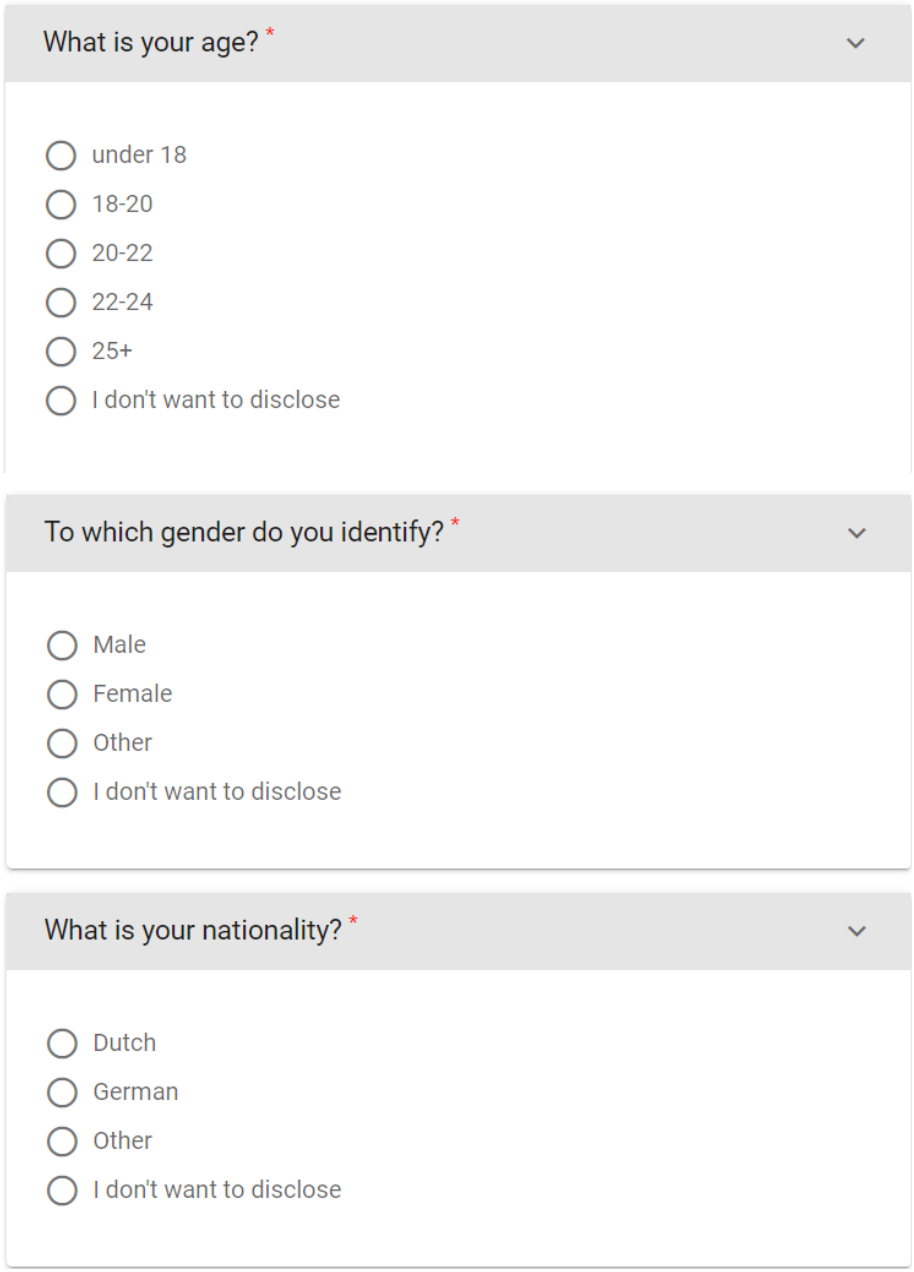
Anonymous data or pseudonyms will be used in any publication. Data that is collected will be stored at a secure location at the University of Twente and on the researchers' secure (encrypted) data carriers. Research data will be stored for a period of 10 years. At the latest after this period has expired, the data will be deleted or anonymized so that they can no longer be traced back to an individual. The research data will only be made available to persons outside the research group if necessary (e.g., for a check on scientific integrity) and only in an anonymous form.
You also have the right to request the researchers to inspect, modify, delete or update your data.

Reimbursement
You will not receive compensation for participating in this study unless you participate in the study through SONA. In that case, you will receive SONA points for your participation.
Contact Information for Questions about Your Rights as a Research Participant
If you have questions about your rights as a research participant or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the Secretary of the Ethics Committee/domain Humanities & Social Sciences of the Faculty of Behavioural, Management and Social Sciences at the University of Twente by ethicscommittee-hss@utwente.nl.

By continuing you agree with the following statements:

- I have read and understood the study information, or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.
- I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.
- I understand that information I provide will be used for data analysis and research into the topic of mental models on university websites.
- I understand that personal information collected about me that can identify me, such as my name or personal details, will not be shared beyond the study team.
- I give permission for the answers that I provide to be archived in anonymized transcripts so they can be used for future research and learning.

Appendix E – Questionnaire Card Sorting



What is your native language? *

- ◯ Dutch
- ◯ German
- ◯ English
- ◯ Other
- ◯ I don't want to disclose

What level of education are you currently following? *

- ◯ VWO
- ◯ First year student
- ◯ Second year student
- ◯ Third year student or higher
- ◯ I don't want to disclose

How often do you use University websites on average? *

- ◯ Daily
- ◯ Weekly
- ◯ Monthly
- ◯ Every 2-3 months
- ◯ Less then 3 times a year
- ◯ I don't want to disclose

How familiar are you with "Card Sorting"? *

- ◯ I used it as a research method myself
- ◯ I never used it but know how it works
- ◯ I have only heard of it
- ◯ I don't know it
- ◯ I don't want to disclose

Appendix F – Instructions Card Sorting

Instructions

Welcome again!

First of all, it is important to know that you can only participate in this study if you are 18 or older. If you are younger than 18 we want to thank you for your enthusiasm but sadly you cannot participate.

On the next page, you will find 34 cards on the left side of the screen. Each card contains an item that can be found on a university website. Your job is to make piles of cards that you feel belong together. In the end, you should name these groups as categories. Keep in mind that there are no right or wrong answers, only your opinion!

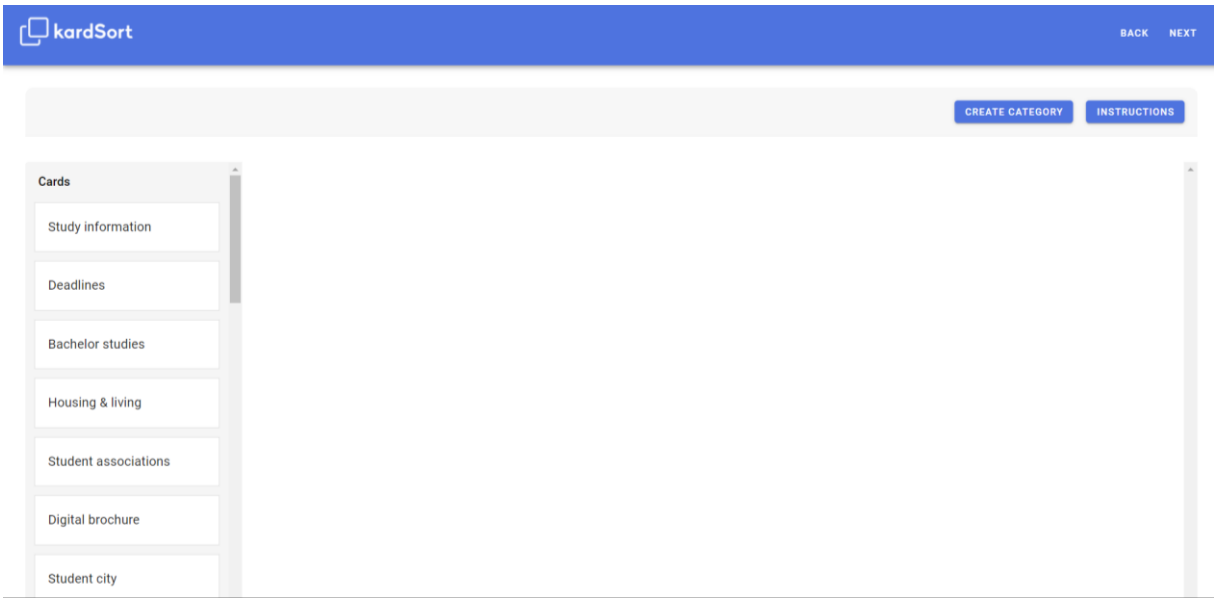
The task works as following:

- To start you need to create a new category using the top right corner button. If you do not know the name of the category you can give it a random name like "1". You are unlimited in the number of categories you want to create.
- To put words in the categories you simply drag them from the left section to the category you feel it belongs. You can also drag cards from one category to the other.
- If you feel that a certain card absolutely does not fit with any of the other cards, you can create a category for just that word.
- When you feel that a specific card belongs to more than one group, please choose the best of those possibilities.
- In the end you need to name the categories in a way that makes sense. You can change the name of the category by selecting the three vertical dots at the right upper corner of every category.
- If, in the end, you have an unused category, you can delete it by using the same three dots.
- You can always find these instruction by selecting "Instructions" in the top right corner.

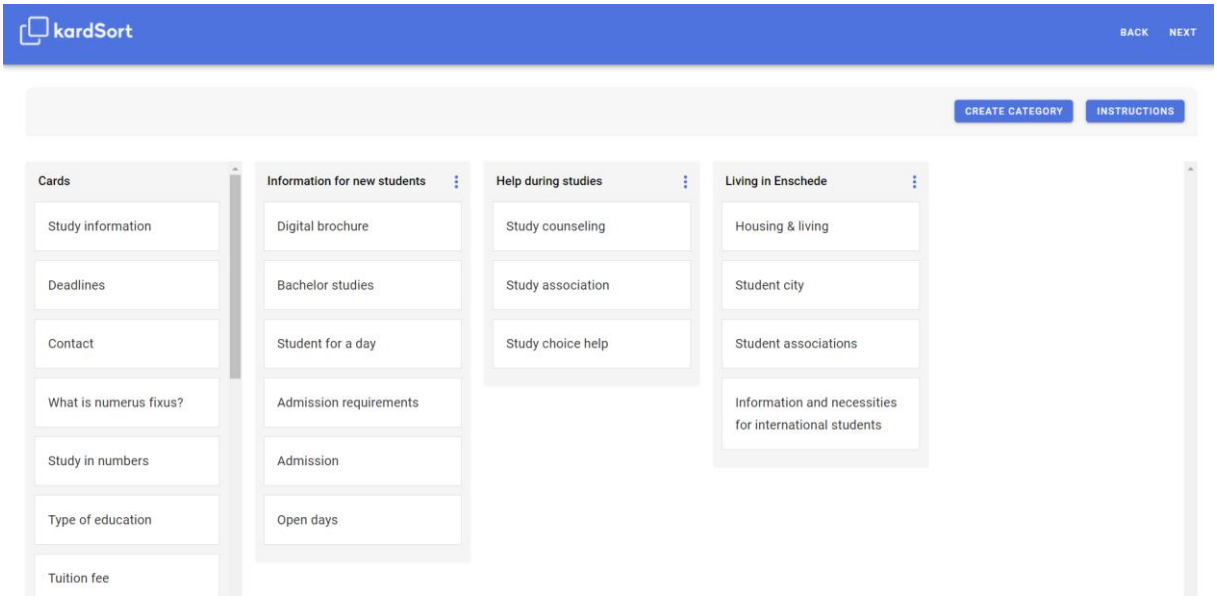
Once again, there are no right or wrong answers!
Enjoy the experience!

Appendix G – Overview of Card Sort Task

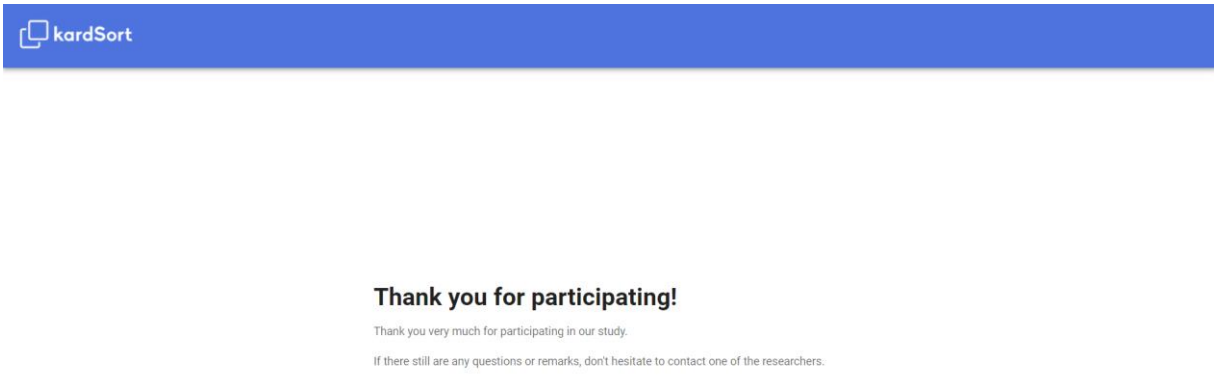
Screenshot of the card sorting task before sorting the cards.



Screenshot of the card sorting task with an example set of sorted cards.



Acknowledgement page after finishing the task.



Appendix H – Item-by-Category Matrix

Item-by-group matrix with suggested category labels from participants. Merged labels are shown that are used assigning category names.

Note. The dark blue lines indicate the clusters; the lighter blue lines indicate subclusters.

Appendix I – Informed Consent Perceived Usability Test

Please read this document carefully before you choose to proceed. You are going to participate in a research study titled "Eliciting Mental Models of University Students Using Card Sorting and the Perceived Usability".

Goal of the study
The goal of this study is to evaluate the perceived usability of websites that are, or are not compatible with where people expect certain information to be placed.

The experiment
You will perform multiple tasks on two websites and will fill out a questionnaire two times about these websites. Your performance on the tasks will not be evaluated, only the perceived usability of the websites. This experiment has been reviewed and approved by the BMS Ethics Committee/domain Humanities & Social Sciences.

Benefits and risks of participating
There are no physical, legal or economic risks associated with your participation in this study.

Voluntary participation and withdrawal
Participation in this study is completely voluntary. As a participant, you may stop your participation in the study at any time or refuse to allow your data to be used for the study, without any reasons.

Data Confidentiality
No confidential information or personal data from or about you will be disclosed in any way that will allow anyone to recognize you. Before the research data is released to the public, your data will be anonymized. The research data will only be made available to persons outside the research group if necessary (e.g., for a check on scientific integrity) and only in an anonymous form.

Contact Information for Questions about Your Rights as a Research Participant
If you have questions about your rights as a research participant or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the Secretary of the Ethics Committee/domain Humanities & Social Sciences of the Faculty of Behavioural, Management and Social Sciences at the University of Twente by ethicscommittee-hss@utwente.nl.

By signing you agree with the following statements:

- I have read and understood the study information, or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.
- I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.
- I understand that information I provide will be used for data analysis and research into the topic of mental models with regard to university websites.
- I understand that personal information collected about me that can identify me, such as my name or personal details, will not be shared beyond the study team.
- I give permission for the answers that I provide to be archived in anonymized transcripts so they can be used for future research and learning.

Name:_____ Date:_____

Signature_____

Appendix J – System Usability Scale

In the following questionnaire please check the box that reflects your immediate response to each statement. Don't think too long about each statement. If you don't know how to respond, simply check the middle box.

	Strongly Disagree				Strongly Agree
I think that I would like to use this website frequently.	○	○	○	○	○
I found the website unnecessarily complex.	○	○	○	○	○
I thought the website was easy to use.	○	○	○	○	○
I think that I would need the support of a technical person to be able to use this website.	○	○	○	○	○
I found the various functions in this website were well integrated.	○	○	○	○	○
I thought there was too much inconsistency in this website.	○	○	○	○	○
I would imagine that most people would learn to use this website very quickly.	○	○	○	○	○
I found the website very cumbersome to use.	○	○	○	○	○
I felt very confident using the website.	○	○	○	○	○
I needed to learn a lot of things before I could get going with this website.	○	○	○	○	○